台灣農業研究 (J. Taiwan Agric. Res.) 72(2):97–112 (2023)
DOI:10.6156/JTAR.202306_72(2).0002 研究報告

雙重酶切 DNA 定序技術應用於番椒種原基因體分析

林思妤[1] 王昭月[1] 陳涵葳[1] 李承彬[1] 林大鈞[2] 杜元凱[1,*]

摘要

林思妤、王昭月、陳涵葳、李承彬、林大鈞、杜元凱。2023。雙重酶切 DNA 定序技術應用於番椒種原基因體分析。台灣農業研究 72(2):97–112。

次世代定序技術已成為作物基因體研究相當常見的工具，而為集中定序資源與降低定序成本，已逐漸發展出多種簡化基因體定序技術，然而不同作物、不同族群組成所適宜的文庫建置方法 (如限制酶選擇或是定序量等) 有差異，若能針對目標作物及族群組成進行方法上的調整，將可使定序資源發揮最大效率。本研究使用簡化基因體技術之一：雙重酶切 DNA 定序技術 (double-digest restriction site-associated DNA sequencing; ddRAD-seq) 針對不同種番椒 (*Capsicum* spp.) 種原基因體進行分析，第一部分比較 *Pst*I/*Msp*I 及 *Pst*I/*Mse*I 限制酶酵素組合，電腦模擬結果顯示兩種組合產生的可被定序片段為總片段之 2.01% 及 0.34%，而實際 2 個酵素組合的定序結果顯示，以 *Pst*I/*Msp*I 組合 (19–28 M) 比 *Pst*I/*Mse*I 組合 (15–17 M) 得到較多成對且比對專一的讀序，因此不論在建庫資源利用上或是成功對應回參考序列的讀序數量/比例上，*Pst*I/*Msp*I 皆較 *Pst*I/*Mse*I 表現佳。本研究第二部分針對 28 個番椒收集系 (涵蓋 5 個番椒馴化種) 以 *Pst*I/*Msp*I 酶切並進行 ddRAD 定序，共得到 14,230 個高品質單一核苷酸多型性 (single nucleotide polymorphism; SNP)，平均分布於番椒 12 對染色體並以染色體兩端基因較密集處較多；利用此些 SNP 進行集群分析，亦能成功將不同種番椒品系正確分類。此說明本研究使用的一系列建庫、定序、生物資訊分析等方法，可用於橫跨不同種的番椒基因體分析，未來將以此為基礎，針對台灣番椒種原建立其基因體資訊，作為台灣番椒育種的根基。

關鍵詞：簡化基因體定序、雙重限制酶、番椒、基因體分析。

前言

番椒 (*Capsicum* spp.) 包含辣椒及甜椒，為全球第二大的茄科作物，具極高經濟價值。其最早起源自中南美洲，為人類利用已有 5,000–6,000 年歷史，目前甜椒主要產區為歐洲及美洲的溫帶及亞熱帶地區，辣椒產區則遍布各大洲熱帶及亞熱帶地區。番椒屬目前已發現至少涵蓋 38 個種，其中 *Capsicum annuum*、*C. baccatum*、*C. chinense*、*C. frutescens* 及 *C. pubescens* 為馴化種，當中又以 *C. annuum* 最為廣泛栽培及應用於育種，且由於長期的人為選拔，在果實大小、形狀、顏色及辣度上皆有非常高的多樣性 (Du *et al*. 2019)。然而面對氣候變遷及人們不斷地求新求變，有必要持續進行番椒品種改良以兼顧栽培及市場偏好。種原為作物改良的基礎，若能瞭解番椒種原的基因體資訊，將有助於加速新品種育成。

目前番椒已有數個全基因體定序資料釋出，Kim *et al*. (2014) 釋出第一個番椒全基因體定序 *C. annuum* ‘CM334’，同年 Qin *et al*. (2014) 釋出 *C. annuum* ‘Zunla-1’ 及 ‘Chiltepin’ (*C. annuum* var. *glabriusculum*) 的定序資料，而後 Kim *et al*. (2017) 針對 *C. chinense* ‘PI159236’ 及 *C. baccatum* ‘PBC81’ 進行解序，Hulse-Kemp *et al*. (2018) 完成 *C. annuum* 一代雜交種 (F_1) 的全基因體定序。此些研究皆顯示番椒基因

投稿日期：2022 年 8 月 12 日；接受日期：2023 年 1 月 7 日。
[*] 通訊作者：yktu@tari.gov.tw
[1] 農委會農業試驗所生物技術組助理研究員。台灣 台中市。
[2] 農委會農業試驗所生物技術組聘用副研究員。台灣 台中市。

體相當大 (約 3−3.5 Gb) 且基因體十分複雜，全基因體 80% 以上由重複序列 (repetitive element) 組成，約具有 35,000 個基因。

全基因體定序所需費用相當高昂，因此西元 2000 年後逐漸發展出簡化基因體定序相關技術，如 reduced representation shotgun sequencing (RRS) (Altshuler *et al*. 2000)、genotyping-by-sequencing (GBS) (Elshire *et al*. 2011) 及 restriction site-associated DNA sequencing (RADseq) (Baird *et al*. 2008) 等技術，上述技術皆是利用定序特定限制酶剪切位置兩側的核酸序列，亦即依據所使用的限制酶，僅定序大約 0.5–4% 的全基因體組核酸序列，因此可大幅降低每個樣本所需定序費用。基於簡化基因體定序概念，後續衍伸出相當多相似的方法，截至 2017 年已至少有 36 篇方法學的發表，現今亦已廣泛應用於多種作物之基因定位、族群結構分析及親緣分析等研究 (Campbell *et al*. 2018)。

目前已有多篇研究利用簡化基因體定序技術進行番椒種原相關研究，如 Han *et al*. (2018) 以 GBS 定序技術進行 208 個番椒品系 (*C. annuum*、*C. chinense* 及 *C. frutescens*) 辣椒素基因座探勘，Ramchiary & Kole (2019) 利用 double-digest RAD (ddRAD) 分析 5 個番椒馴化種 (186 個品系) 之果實方向性，Pereira-Dias *et al*. (2019) 以 GBS 分析 190 個番椒品系 (橫跨至少 5 個種) 之遺傳歧異度及族群結構，Potnis *et al*. (2019) 則以 GBS 為基礎進行 287 個美國農部番椒種原抗細菌性斑點病 (*Xanthomonas gardneri*) 基因座定位。進行簡化基因體定序時，限制酶的選擇相當重要，會影響到酶切後片段數目及大小，從而影響到讀序 (reads) 及單一核苷酸多型性 (single nucleotide polymorphism; SNP) 的數量及分布。綜上各篇利用簡化基因體定序技術於番椒種原相關研究文獻，目前常見以 1 個限制酶 (*Ape*KI) 酶切 (Nimmakayala *et al*. 2016; Taitano *et al*. 2019)，亦有用 2 個限制酶酶切者，如 *Pst*I/*Msp*I (Yamazaki *et al*. 2020) 或 *Pst*I/*Mse*I 組合 (Han *et al*. 2018)。

本研究團隊將針對台灣種原庫中的 4,800 多份番椒種原進行基因體研究，並以基因型資料建立核心種原，由於番椒種原數量龐大，且基因體亦相當龐大 (3–3.5 Gb)，因此有必要進行先期試驗，以找到定序資源利用效率最佳的方法，並初步對番椒種原進行遺傳背景分析。本研究分為 2 部分，第一部分以電腦模擬 (*in silico*) 及活體外試驗 (*in vitro*) 評估 2 種番椒常見的限制酶酵素組合—*Pst*I/*Msp*I 及 *Pst*I/*Mse*I 用於 ddRAD 中的效率，第二部分則對 28 個番椒收集系進行遺傳歧異度的分析，以對番椒種原之遺傳背景有初步瞭解。

材料與方法

植物材料及 DNA 萃取

本研究番椒品系來自行政院農業委員會農業試驗所之國家作物種原中心種原庫及育種材料，涵蓋 5 個番椒馴化種，包含 17 個 *C. annuum*、3 個 *C. baccatum*、3 個 *C. chinense*、2 個 *C. frutescens*、2 個 *C. pubescens* 及 1 個未知種 *C.* sp.，合計共 28 個收集系 (表 1)。28 個收集系中有 10 個收集系屬自交系 (名稱為 SP 開頭)，經自交純化 4–6 代，其他 18 個收集系則來自種原庫，未經自交純化。此些番椒品系於 2020 年 10 月栽培於農業試驗所溫網室 (位於台中霧峰)，採離土栽培，肥培管理依照慣行農法進行。

每個品系採摘同一單株 2 片嫩葉冷凍乾燥備用，再以 DNeasy Plant Mini kit (Qiagen GmbH, Hilden, Germany) 進行 DNA 萃取，最後以 Qubit dsDNA BR 試劑及 Qubit 螢光儀 (Thermo Fisher Scientific, Waltham, MA, USA) 進行 DNA 定量，並以 1% agarose gel 確認 DNA 品質。

雙重限制酶組合評估

ddRAD-seq 概念為使用 1 稀疏切位 (rare cutter) 以及 1 常用切位 (common cutter) 之限制酶剪切作物基因體核酸，以獲取合適數目與大小基因體核酸片段進行定序，進而達到簡化基因組定序效果 (Elshire *et al*. 2011)。考量試驗進行成本 (限制酶用量、建庫及定序費用)，又相關研究指出使用甲基化敏感之限制酶較易於染色體富含基因區域進行剪切 (Han *et al*. 2018)，綜合上述因素以及常用在番椒 ddRAD-

表 1. 參試番椒樣本清單。

Table 1. Sample list of *Capsicum* spp.

ID	Species	Name	Heterozygous proportion
P0133	*C. annuum*	01A00493	0.079
P0134	*C. annuum*	01A00494	0.101
P0164	*C. annuum*	05A01225	0.106
P0287	*C. annuum*	97A05512	0.084
P0565	*C. annuum* var. *aviculare*	95A04169	0.116
P1025	*C. annuum*	SP20	0.128
P1027	*C. annuum*	SP40	0.110
P1028	*C. annuum*	SP41	0.071
P1032	*C. annuum*	SP63	0.110
P1033	*C. annuum*	SP71	0.140
P1034	*C. annuum*	SP80	0.073
P1035	*C. annuum*	SP96	0.076
P1036	*C. annuum*	SP98	0.106
P1037	*C. annuum*	SP100	0.153
P1038	*C. annuum*	SP104Y	0.081
P1044	*C. annuum*	CM334-1-22	0.094
P1050	*C. annuum*	Jalapeno (OP)	0.082
P0357	*C. baccatum* var. *pendulum*	00A00660	0.160
P0445	*C. baccatum* var. *praetermissum*	93A01929	0.183
P0583	*C. baccatum* var. *baccatum*	98A00685	0.102
P0011	*C. chinense*		0.096
P0653	*C. chinense*	94A05917	0.091
P1049	*C. chinense*		0.136
P0033	*C. frutescens*	00A05011	0.114
P0157	*C. frutescens*	04A00477	0.120
P0405	*C. pubescens*	04A02777	0.238
P0406	*C. pubescens*	04A02778	0.112
P0100	*C.* sp.	01A00456	0.082

seq 相關研究 (Zhang *et al*. 2010; Yamazaki *et al*. 2020)，本試驗共評估 2 種限制酶酵素組合 *Pst*I/*Msp*I 及 *Pst*I/*Mse*I，其中識別 6 個核苷酸的限制酶*Pst*I 為 rare-cutter，另一端則以識別 4 個核苷酸的限制酶 *Msp*I 或 *Mse*I 作為 common-cutter。兩種酵素組合皆以電腦模擬及活體外試驗進行評估比較，電腦模擬部分以 R 軟體 SimRAD (Lepais & Weir 2014) 進行，番椒參考基因體採用 CM334 品種 (Kim *et al*. 2014) v.1.6.0 (下載自 http://peppergenome.snu.ac.kr)。

ddRAD 文庫製備及定序

文庫製備依照Peterson *et al*. (2012) 及Kao (2016) 的方法，流程簡述如下：

(1) 雙重限制酶酶切 (digestion)：取 300 ng 的 gDNA，以 *Pst*I-HF®/*Msp*I 或 *Pst*I-HF®/*Mse*I (New England Biolabs, Ipswich, MA, USA) 於 37℃下進行酶切至少 4 h。

(2) 黏合銜接子 (adapter ligation)：將酶切完的 DNA 溶液加入 4 pmol P1 (針對 rare-cutter)、40 pmol P2 (針對 common-cutter) 銜

接子 (P1 及 P2 銜接子序列如表 2)，並以 6.6 Weiss units T4 DNA ligase (New England Biolabs, Ipswich, MA, USA) 及 1 mM rATP (Promega, Madison, WI) 於 16℃反應過夜，之後以 65℃ 20 min 使連接酶 (ligase) 失去活性。

(3) 純化 DNA：以 0.8× Ampure XP beads (Beckman Coulter Genomics, Brea, CA, USA) 去除 buffer 與酵素，並回溶至 42 µL 二次水。

(4) 連接標誌 (index) 及晶片 (flowcell) 連接序列：將帶有 Nextera-XT-Index-Kit2 (Nextera-XT-Index 序列如表 3 所示) 及 flowcell 黏合序列的引子 (0.5 µM)，以 0.02 U Phusion High-Fiedlity DNA Polymerases (Thermo Fisher Scientific, Waltham, MA, USA)、200 µM dNTPs 進行聚合酶連鎖反應，反應條件為 98℃ 30 s、98℃ 10 s、72℃ 30 s，18 cycles；72℃ 10 min；保存於 4℃。

(5) 聚合酶連鎖反應 (polymerase chain reaction; PCR) 產物純化並定量：同樣以 0.8× Ampure XP beads (Beckman Coulter Genomics, Brea, CA, USA) 去除 buffer 與酵素，回溶至 22 µL 二次水後以 qubit (Thermo Fisher Scientific, Waltham, MA, USA) 定量。

(6) 混合樣本並進行產物片段篩選：將不同 index 的 PCR 產物等體積混合，以 BluePippin 系統 (Sage Scientific, Beverly, MA, USA) 篩選 250–600 bp 的片段。

(7) 濃縮文庫並確認品質：以 1× Ampure XP beads (Beckman Coulter Genomics, Brea, CA, USA) 濃縮 DNA 文庫，濃度至少達 10 ng μL^{-1}，NanoDrop 品質檢測需達 260/280 = 1.8–1.9、260/230 > 1.8。

(8) 定序：以 HiSeq X™ Ten (Illumina, San Diego, CA, USA) 次世代定序儀進行定序，採 150 bp pair-end。

定序資料分析及 SNP 探勘

原始定序資料首先使用 FastQC 及 MultiQC 進行讀序品質及數量總結，接續以 Stacks v2.61 軟體的 process_radtag 程式移除品質不佳的讀序，如酵素切位不合、index 序列錯誤、讀序品質 Phred quality score 小於 20 者，再以 BWA-MEM v0.17.7 程式 (Li 2013) 將讀序對應至參考序列 *C. annuum* CM334 v1.60 組裝版本 (http://peppergenome.snu.ac.kr)，使用指令為 bwa mem -M {CM334_reference.fasta} {cleaned_forwaed.fastq} {cleaned_reverse.fastq} > {aligned_SAM}，並經 Picard v2.25.2 軟體 (https://broadinstitute.github.io/picard) 的 SortSam 與 MarkDuplicates 指令篩選可成對對應到參考序列 (properly paired) 且只對應到 1 個基因體上的位置 (uniquely mapped) 的成對讀序，使用指令為 SortSam I = {input} O = {output} SORT_ORDER = coordinate；MarkDuplicates I = {input} O={output} M = {output.metrics}。上述之成對讀序經 samtools v1.7 軟體 (Danecek *et al*. 2021) 的 view 指令確認讀序數量，指令為 samtools view -c -f 2 -F 2308 {input})，並使用 flagstat 或 mpileup 指令分析 BAM 檔案資料完後 (指令為 samtools flagstat {input}；samtools mpileup {input})，進入下一個步驟：SNP 探勘。SNP 探勘以 GATK v4.2.0.0 軟體 (https://gatk.broadinstitute.org/hc/en-us) 的 HaplotypeCaller 程式進行，先產生包含所有基因位點的 GVCF 中間檔案 (指令為 HaplotypeCaller -R {CM334_reference.fasta} -I {input} -O {output} -ERC GVCF")，再以 GenomicsDBImport 及 GenotypeGVCFs 指令將個別樣本的變異位點數據做統整、分析，建立樣本族群變異位點資料庫 (指令為 GenomicsDBImport --genomicsdb-workspace-path {output_path} --sample-name-map {sample_

表 2. 本研究使用之 P1 及 P2 銜接子序列。

Table 2. The nucleotide sequences of the P1 and P2 adapters used in this study.

Name	Description	Sequence
Adapter-P1	Rare-cutter	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCA-3'
Adapter-P2	Common-cutter	5'-CGAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3'

表 3.　本研究使用之 Nextera-XT-Index 序列。

Table 3.　The nucleotide sequences of the Nextera-XT-Index used in this study.

		I7_Index		I5_Index	
Sample	Species	ID	Sequence	ID	Sequence
20F-P0011	*Capsicum chinense*	N701	TAAGGCGA	S513	CTAGTCGA
20F-P0033	*C. frutescens*	N702	CGTACTAG	S513	CTAGTCGA
20F-P0100	*C.* sp.	N703	AGGCAGAA	S513	CTAGTCGA
20F-P0133	*C. annuum*	N704	TCCTGAGC	S513	CTAGTCGA
20F-P0134	*C. annuum*	N701	TAAGGCGA	S515	AGCTAGAA
20F-P0157	*C. frutescens*	N702	CGTACTAG	S515	AGCTAGAA
20F-P0164	*C. annuum*	N703	AGGCAGAA	S515	AGCTAGAA
20F-P0287	*C. annuum*	N704	TCCTGAGC	S515	AGCTAGAA
20F-P0357	*C. baccatum*	N701	TAAGGCGA	S516	ACTCTAGG
20F-P0405	*C. pubescens*	N702	CGTACTAG	S516	ACTCTAGG
20F-P0406	*C. pubescens*	N703	AGGCAGAA	S516	ACTCTAGG
20F-P0445	*C. baccatum*	N701	TAAGGCGA	S517	TCTTACGC
20F-P0565	*C. annuum*	N702	CGTACTAG	S517	TCTTACGC
20F-P0583	*C. baccatum*	N703	AGGCAGAA	S517	TCTTACGC
20F-P0653	*C. chinense*	N704	TCCTGAGC	S517	TCTTACGC
20F-P1025	*C. annuum*	N701	TAAGGCGA	S518	CTTAATAG
20F-P1027	*C. annuum*	N702	CGTACTAG	S518	CTTAATAG
20F-P1028	*C. annuum*	N703	AGGCAGAA	S518	CTTAATAG
20F-P1032	*C. annuum*	N701	TAAGGCGA	S520	ATAGCCTT
20F-P1033	*C. annuum*	N702	CGTACTAG	S520	ATAGCCTT
20F-P1034	*C. annuum*	N703	AGGCAGAA	S520	ATAGCCTT
20F-P1035	*C. annuum*	N704	TCCTGAGC	S520	ATAGCCTT
20F-P1036	*C. annuum*	N701	TAAGGCGA	S521	TAAGGCTC
20F-P1037	*C. annuum*	N702	CGTACTAG	S521	TAAGGCTC
20F-P1038	*C. annuum*	N703	AGGCAGAA	S521	TAAGGCTC
20F-P1044	*C. annuum*	N701	TAAGGCGA	S522	TCGCATAA
20F-P1049	*C. chinense*	N702	CGTACTAG	S522	TCGCATAA
20F-P1050	*C. annuum*	N703	AGGCAGAA	S522	TCGCATAA
20F-P0033-mse	*C. frutescens*	N705	GGACTCCT	S513	CTAGTCGA
20F-P0434-mse	*C. annuum*	N705	GGACTCCT	S515	AGCTAGAA
20F-P1025-mse	*C. annuum*	N705	GGACTCCT	S516	ACTCTAGG
20F-P1028-mse	*C. annuum*	N705	GGACTCCT	S517	TCTTACGC

name_list.txt} -L {params.intervals} --reader-threads 20；GenotypeGVCFs 則為 -R {CM334_reference.fasta} -V gendb://{input.chr_file} -O {output})，接著以 Picard 軟體的 GatherVcfs 指令將分析數據整併成單一 VCF 檔案 (指令為 GatherVcfs I = {input.ch01} I = {input.ch02} I = {input.ch03} I = {input.ch04} I = {input.ch05} I = {input.ch06} I = {input.ch07} I = {input.ch08} I = {input.ch09} I = {input.ch10} I = {input.ch11} I = {input.ch12} O = {output})。最後使用 GATK 軟體的 SelectVariants 指令 (-R {CM334_reference.fasta } -V

{input} --select-type-to-include SNP -O {output_SNP.vcf}) 擷取 SNP 資料。獲得之 SNP 資料以 VCFtools v0.1.15 軟體 (Danecek *et al*. 2011) 篩選 SNP，設定參數為 --minDP 5 --minGQ 15 --max-missing 0.5，篩選後 VCF 檔案匯入 TASSEL v5.0 軟體 (https://www.maizegenetics.net/tassel) 進行總結。過程中以 Integrative Genomics Viewer (Thorvaldsdóttir *et al*. 2013) 及 R 語言 CMplot 套件 (https://CRAN.R-project.org/package=CMplot) 輔助讀序及 SNP 視覺化呈現。

主成分分析與親緣分析

主成分分析及親緣分析皆於 TASSEL5 進行。主成分分析設定最低維度參數為 2，搭配共變異數矩陣 (covariance matrix) 進行分析，而後將獲得的 PC1 以及 PC2 座標匯出為 tsv 格式，後續則使用 Past v4.03 軟體 (Hammer *et al*. 2001) 進行視覺化呈現，並標註每個樣本的對應位置。親緣分析則依據主成分分析 (principal component analysis; PCA) 的輸出資料，搭配 Past v4.03 軟體的聚類分析 (clustering) 功能繪製成樹狀圖，調整參數為非加權算術平均連鎖法 (unweighted pair-group method using the average approach; UPGMA) 進行相似度計算。

結果

比較不同限制酶組合用於番椒 ddRAD 定序

為選擇較適合番椒 ddRAD 的限制酶組合，本研究比較 2 種番椒簡化基因體定序之酵素組合—*Pst*I/*Msp*I 及 *Pst*I/*Mse*I。電腦模擬結果顯示，*Pst*I/*Msp*I 組合可產生 322,050 個片段，*Pst*I/*Mse*I 組合可產生 5,509,217 個片段，後者產生的片段數為前者的 17 倍，但多屬小片段 (95% 片段短於 403 bp)，而前者產生的片段長度則涵蓋較廣，95% 片段介於 0–7,467 bp (表 4)。而針對目標 150–500 bp 間的片段，*Pst*I/*Msp*I 組合及 *Pst*I/*Mse*I 組合分別產生為 54,209 及 1,268,219 個，前者片段大小分布較均勻，而後者有短片段較多的趨勢。再更深入針對目標 150–500 bp 間片段探討酵素酶切型態，發現片段兩端為不同酵素酶切者 (AB + BA，AB 代表 5′ 端以酵素酶 *Pst*I 切、3′ 端以酵素酶 *Msp*I 或 *Mse*I 切之讀序；BA 代表 5′ 端以酵素酶 *Msp*I 或 *Mse*I 切、3′ 端以酵素酶 *Pst*I 切之讀序)，*Pst*I/*Msp*I 組合及 *Pst*I/*Mse*I 組合各有 13,242 及 39,110 個，分別占全部 150–500 bp 目標片段之 24.4% 及 3.1% (圖 1)。片段兩端為相同酵素酶切的片段 (AA + BB，AA 代表兩端皆為酵素酶 *Pst*I 切之讀序；BB 代表兩端皆為酵素酶 *Msp*I 或 *Mse*I 切之讀序) 及 BA 形式酶切的片段，在 PCR 擴增階段因為 Y 型 adapter 的設計而不會被擴增，所以只有 AB 的片段可被擴增從而被定序，因此依據電腦模擬結果，*Pst*I/*Msp*I 組合及 *Pst*I/*Mse*I 組合分別得到 6,484 及 18,926 個可被定序的片段，分別為總片段之 2.01% 及 0.34%。

實際以 3 個樣本 (P0033、P1025、P1028) 觀察 *Pst*I/*Msp*I 及 *Pst*I/*Mse*I 組合之定序結果 (表 5)，*Pst*I/*Msp*I 組合比 *Pst*I/*Mse*I 組合得到較少讀序，前者為 20–32 M 讀序，後者為 35–38M 讀序，但經讀序品質過濾後，成對對應到參考序列 (properly paired) 且只對應到 1 個基因體上的位置 (uniquely mapped) 的讀序數量，反而 *Pst*I/*Msp*I 組合比 *Pst*I/*Mse*I 組合得到較多

表 4. 電腦模擬不同酵素組合對番椒參考基因體 (*Capsicum annuum* 'CM334') 酶切所得到之讀序數量。

Table 4. *In silico* simulation of reads number of pepper reference genome (*Capsicum annuum* 'CM334') digested with different restriction enzyme pairs.

Restriction enzyme pair	Total reads number	bp of 95% reads	Number of reads in 150–500 bp	Number of reads in 150–500 bp, AA + BB type	Number of reads in 150–500 bp, AB + BA type	Number of reads in 150–500 bp, AB type
*Pst*I/*Msp*I	322,050	7,467	54,209	40,967	13,242	6,484
*Pst*I/*Mse*I	5,509,217	403	1,268,219	1,229,109	39,110	18,926

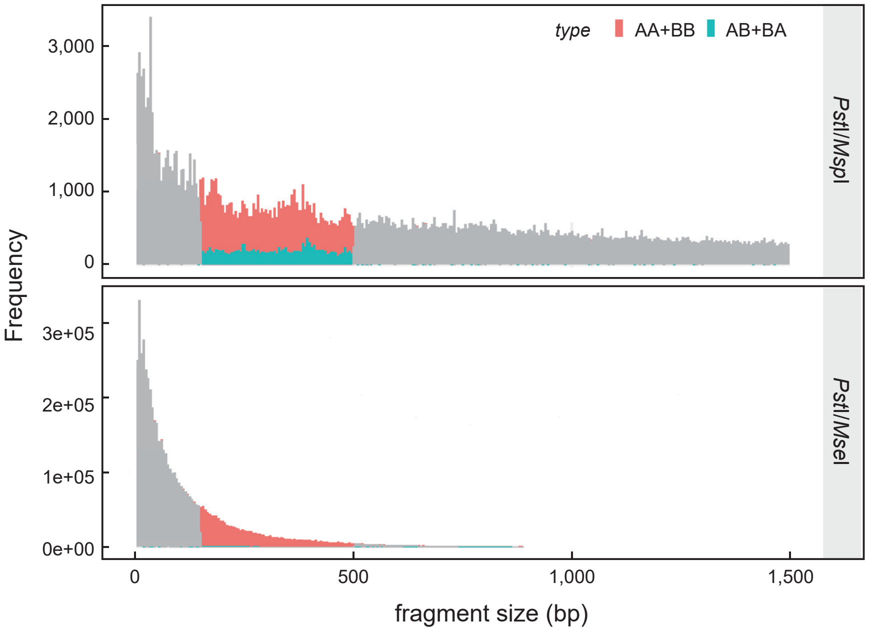

圖 1. 電腦模擬不同酵素組合對番椒參考基因體 (*Capsicum annuum* 'CM334') 酶切所得到之讀序分布情形。灰色區域表示酶切後所有讀序，紅色區域為長度介於 150–500 bp 且兩端為相同酵素酶切 (AA + BB) 之讀序，藍色區域為長度介於 150–500 bp 且兩端為不同酵素酶切 (AB + BA) 之讀序。AA：兩端皆為酵素酶 *Pst*I 切之讀序，BB：兩端皆為酵素酶 *Msp*I 或 *Mse*I 切之讀序，AB：5′ 端以酵素酶 *Pst*I 切、3′ 端以酵素酶 *Msp*I 或 *Mse*I 切之讀序，BA：5′ 端以 *Msp*I 或 *Mse*I 切、3′ 端以酵素酶 *Pst*I 切之讀序。

Fig. 1. *In silico* simulation of the read distribution came from digesting the *Capsicum annuum* reference genome ('CM334') with different pairs of restriction enzymes. Grey area: all the restriction fragments; red area: fragments in the range 150–500 bp and digested with the same restriction enzymes in both end; blue area: fragments in the range 150–500 bp and digested with different restriction enzymes in both end. AA: fragments digested with the same restriction enzymes *Pst*I in both end; BB: fragments digested with the same restriction enzymes *Msp*I or *Mse*I in both end; AB: fragments digested with restriction enzyme *Pst*I in 5′ end and restriction enzyme *Msp*I or *Mse*I in 3′ end; BA: fragments digested with restriction enzyme *Msp*I or *Mse*I in 5-end and restriction enzyme *Pst*I in 3′ end.

讀序，分別約為 19–28 M 及 15–17 M 條讀序，高品質讀序數量約為原始讀序數量的 87−92% 及 43−46%。比較兩個組合之讀序分布情形，由圖 2 可見，兩個組合之讀序皆涵蓋番椒 12 條染色體，且皆有讀序在染色體兩端較多的情形，其中 *Pst*I/*Mse*I 組合的讀序於各區域數量上落差較小，*Pst*I/*Msp*I 組合的讀序則在部分區域有讀序較多的情形。

番椒種原基因體分析

本研究第二部分針對 28 個番椒收集系以 *Pst*I/*Msp*I 酶切並進行 ddRAD 定序，平均每個樣本得到 26 M 條讀序，拼貼回番椒參考序列的總長度為 137,970,405 bp，約占全基因體 3.9%。經讀序品質篩選及探勘 SNP 等步驟後，共得到 850,400 個 SNP，其中深度達 5 以上、正確率 97% 以上、遺失率小於 50% 的高品質 SNP 共 14,230 個 (圖 3)。此些 SNP 涵蓋番椒 12 條染色體，平均每條染色體有 1,185 個 SNP，以第 1 條染色體有最多 SNP，第 8 條染色體有最少 SNP；12 條染色體平均 204 Kb 有 1 個 SNP，密度最高為第 2 條染色體，每 131 Kb 出現 1 個 SNP，最低為第 10 條染色體，約 256 Kb 才會出現 1 個 SNP (表 6)。

28 個番椒收集系以種的位階進行分類 (*C. annuum*、*C. baccatum*、*C. chinense*、*C. frutescens*、*C. pubescens* 及未知種 *C.* sp.) 並計算不同種番椒成對對應到參考序列且只對應到 1 個基因體上的位置的讀序數量，不同種番椒材料經 *Pst*I/*Msp*I 酶切並進行 ddRAD 定序

表 5. 不同酵素組合酶切得到的讀序數量。
Table 5. Number of sequencing reads digested with different restriction enzyme pairs.

Enzyme pairs	Line	Total reads	Mapped reads	Uniquely mapped reads	Uniquely mapped and properly paired reads
*Pst*I-*Msp*I	P0033	25,729,102	23,606,978	23,398,772	22,362,212
		100%	92%	91%	87%
	P1025	32,647,308	29,575,297	29,421,077	28,613,694
		100%	91%	90%	88%
	P1028	20,797,859	19,842,359	19,717,124	19,204,094
		100%	95%	94%	92%
*Pst*I-*Mse*I	P0033	35,753,621	28,320,981	23,979,588	15,586,808
		100%	79%	67%	43%
	P1025	37,621,037	29,687,317	25,363,306	16,913,924
		100%	79%	67%	45%
	P1028	37,924,974	30,312,547	25,920,205	17,507,400
		100%	80%	68%	46%

所得可進行後續 SNP 分析之讀序數目約占原始讀序數目之 75.6–93.9% (圖 4)。

28 個番椒收集系來自 5 個番椒馴化種，異質度 (heterozygous proportion) 介於7.1–23.8%，平均為 11.2% (表 1)。觀察不同種樣本的異質度，並無顯著偏高或偏低的情況；再深入觀察自交系 (名稱為 SP 開頭) 與種原 (名稱為作物種原庫編碼 xxAxxxxx) 的異質度，亦無顯著何者較高或較低的情況。

番椒種原集群分析

以 28 個番椒收集系之 14,230 個 SNP 進行主成分分析，結果顯示約可分為 4 群，其中 *C. baccatum* 及 *C. pubescens* 可各自清楚地被劃分為一群，分別位於圖 5 左下角及左上角，17 個 *C. annuum* 品系及 1 個未知品系 (*C.* sp.) 則皆位於圖 5 右側且相當集中，*C. chinense* 及 *C. frutescens* 於圖正中央被劃分為同一群，不過有 1 個 *C. chinense* 及 1 個 *C. frutescens* 也被歸類至 *C. annuum* 一群。

非加權算術平均連鎖法 (UPGMA) 的結果 (圖 6) 顯示，*C. pubescens* 及 *C. baccatum* 的品系最早被分開，分別為 Group I 及 Group II，*C. annuum*、*C. chinense*、*C. frutescens* 及 sp. 的品系位於同一分支 (Group III)，此分支若再細分，*C. chinense* 及 *C. frutescens* 的部分品系可再區分為一分支 Group III-a，*C. annuum* 及 *C.* sp. 則全數落於 Group III-b。此結果與前述主成分分析結果一致。

討論

簡化基因體定序利用限制酶酵素剪切或物理震盪方式，將定序資源集中少部分基因體，因此在經費有限的前提下，對於基因體較大的作物，簡化基因體定序為相當合適的定序方法。番椒的基因體相當大 (3–3.5 Gb)，為同屬茄科的番茄 (950 Mb)、馬鈴薯 (844 Mb) 、茄子 (1.13 Gb) 基因體的 3 倍以上，且其基因體組成亦相當複雜，約 75–80% 由 repetitive elements 組成 (Hulse-Kemp *et al*. 2018)。有鑑於此，本研究擬以簡化基因體定序方法進行番椒種原基因體資訊分析，以降低定序成本，且相比於常見使用 1 個酵素酶切或物理剪切方式，本研究採用 2 個酵素酶切，使得在相同定序通量下，定序片段更加集中部分區域，提升基因體資料的正確率 (Peterson *et al*. 2012)。

酵素的選擇會影響到基因體的二次抽樣 (subsampling)，導致酶切後片段數目、大小不同，從而影響到讀序及 SNP 的數量及分布 (Cumer *et al*. 2021)。在本研究中，第一部分探討常見用於番椒的雙酵素酶切組合 *Pst*I/*Msp*I

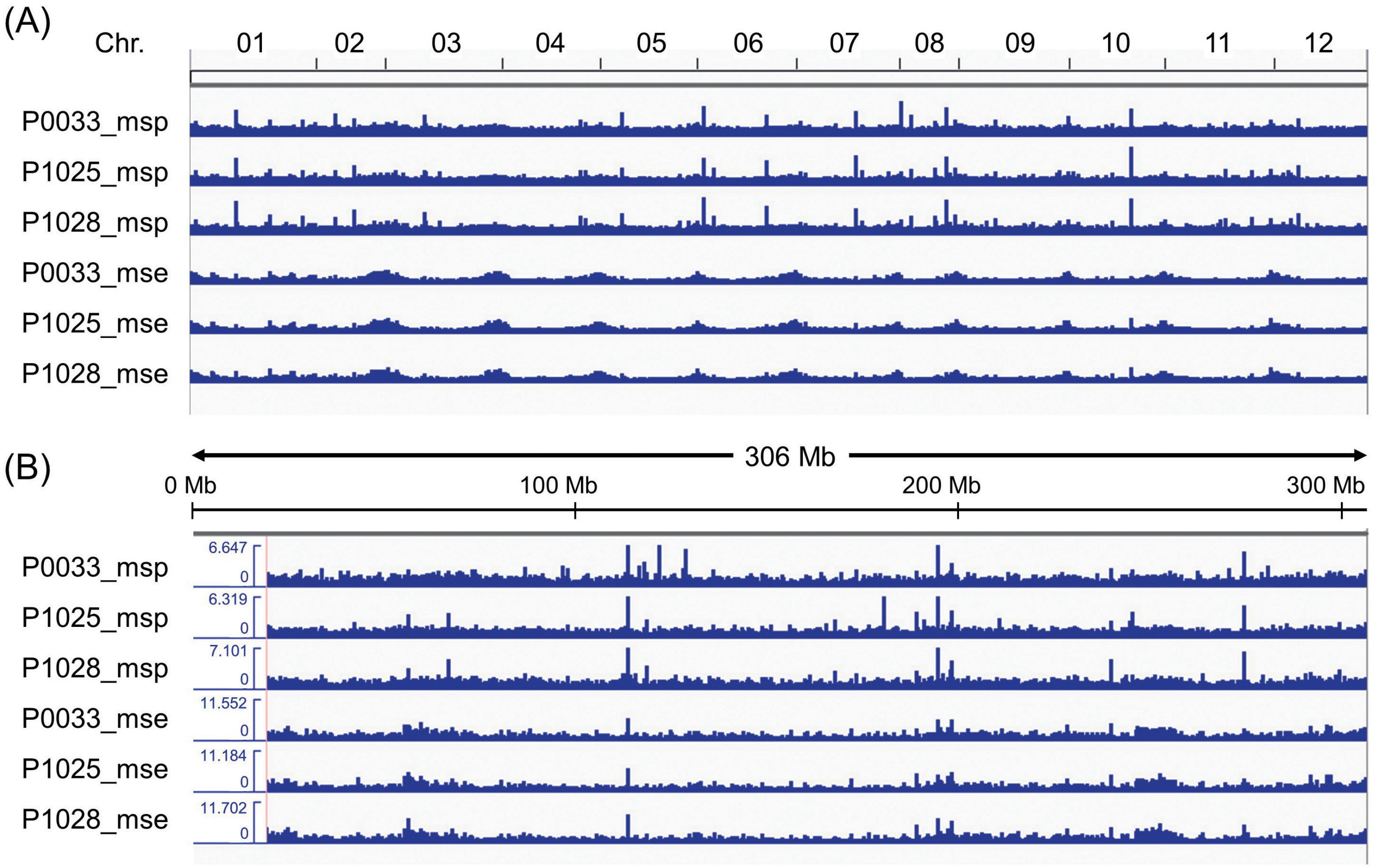

圖 2. 不同酵素組合針對相同樣本酶切所得到讀序之深度與分布概況。(A) 番椒 12 條染色體；(B) 番椒第一條染色體。X 軸代表染色體位置，Y 軸代表該區域之深度讀值。

Fig. 2. Distribution of the read depths with 6 restriction enzymes and pepper varieties combinations (A) along 12 chromosomes and (B) on chromosome 1. X axis represents the location of chromosome, and Y axis represents the reads depth of the region.

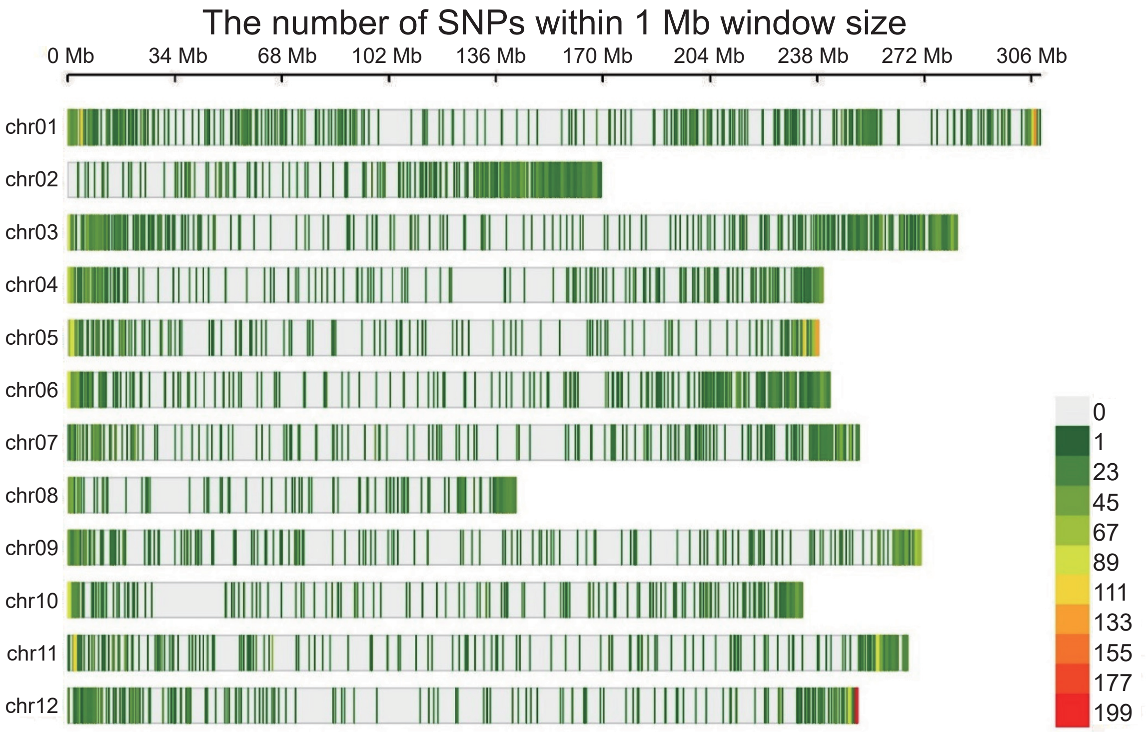

圖 3. 28 個番椒收集系的 14,230 個高品質單一核苷酸多型性位點分布。位點是使用 *Pst*I/*Msp*I 雙酶切的簡化基因體定序技術所產生。

Fig. 3. Distribution of 14,230 high-quality single nucleotide polymorphisms (SNPs) derived from the *Pst*I/*Msp*I double-digested RADseq technique for 28 pepper accessions.

表 6. 平均每條染色體 SNP 數量及密度。

Table 6. Number and density of single nucleotide polymorphism (SNP) per chromosome.

Chromosome	Number of SNP	Average distance (bp) between SNP
1	1,840	167,990
2	1,296	130,830
3	1,685	167,822
4	999	240,361
5	1,099	216,875
6	1,353	179,040
7	1,075	233,761
8	582	244,616
9	1,089	248,928
10	910	256,398
11	1,130	236,168
12	1,172	214,104
Average	1,186	203,655

及 *Pst*I/*Mse*I，何者較適宜用於番椒 ddRAD 定序。電腦模擬結果顯示，*Pst*I/*Mse*I 產生的片段數雖為 *Pst*I/*Msp*I 的 17 倍，但大部分皆為短片段，在目標片段選拔的階段會被移除高比例的片段，較浪費建庫資源；從可被定序片段的數量來看，*Pst*I/*Msp*I 及 *Pst*I/*Mse*I 分別得到 6,484 及 18,926 個可被定序的片段，為總片段之 2.01% 及 0.34%，此亦顯示 *Pst*I/*Msp*I 組合在建庫過程中資源較不浪費。

以實際 2 個酵素組合的定序結果來看，得到的讀序數量 *Pst*I/*Mse*I 多於 *Pst*I/*Msp*I 組合，與預期 (電腦模擬) 結果一致，但經過對應回參考序列 (alignment) 的步驟後，成對對應到參考序列 (properly paired) 且只對應到 1 個基因體上的位置 (uniquely mapped) 的讀序數量，卻以 *Pst*I/*Msp*I 組合 (19–28 M) 比 *Pst*I/*Mse*I 組合 (15–17 M) 得到較多讀序，分別約

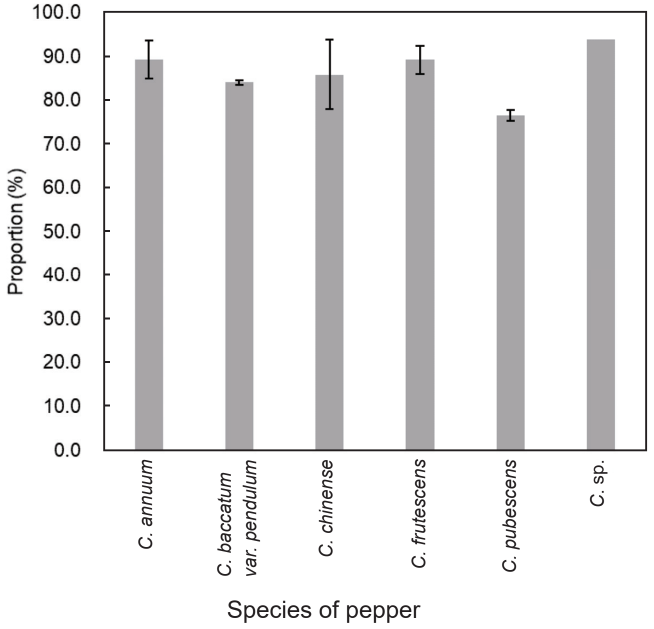

圖 4. 28 個番椒收集系以種的位階進行分類 (*Capsicum annuum*、*C. baccatum*、*C. chinense*、*C. frutescens*、*C. pubescens* 及未知種 *C.* sp.)，計算不同種番椒材料成對對應到參考序列 (properly paired) 且只對應到 1 個基因體上的位置 (uniquely mapped) 的讀序數量。誤差線為標準差 (*C.* sp. 僅有 1 個樣本故未計算與標示標準差)。

Fig. 4. 28 accessions were classified by *Capsicum annuum*, *C. baccatum*, *C. chinense*, *C. frutescens*, *C. pubescens* and *C.* sp. into 6 groups. The corresponding properly paired and uniquely mapped reads for each group were calculated. Error bar: standard deviation (SD) (Only 1 sample of *C.* sp. was included, no SD for *C.* sp. was calculated and labeled).

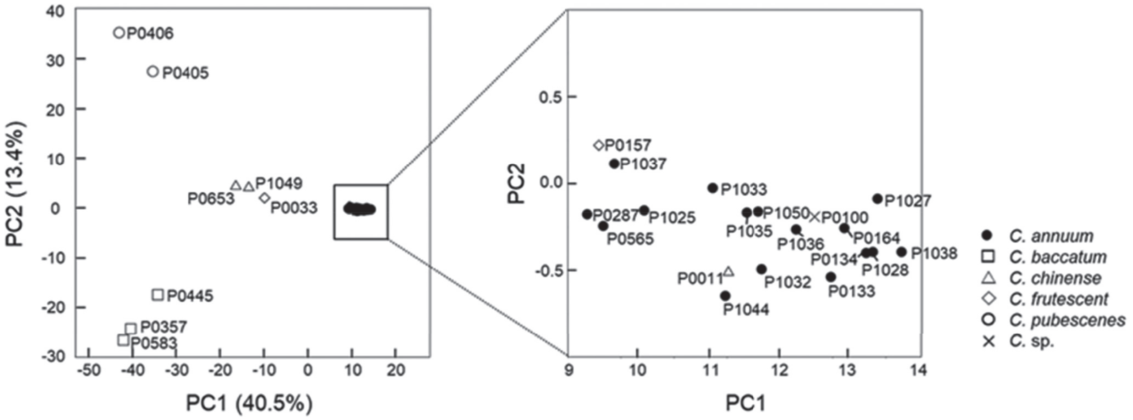

圖 5. 28 個番椒收集系之主成分分析圖。以 14,230 個 SNP 資料所計算，每個點表示不同品系，點的顏色表示不同種。X 軸與 Y 軸分別代表 PC1 (40.5%) 與 PC2 (13.4%)。

Fig. 5. The principal component analysis (PCA) plot of 28 pepper accessions using genotypes of 14,230 single nucleotide polymorphisms (SNPs). Each point represents the SNP profile of an individual. The color represents *Capsicum* species. X axis represents the variance explained by PC1 (40.5%), and Y axis represents the variance explained by PC2 (13.4%).

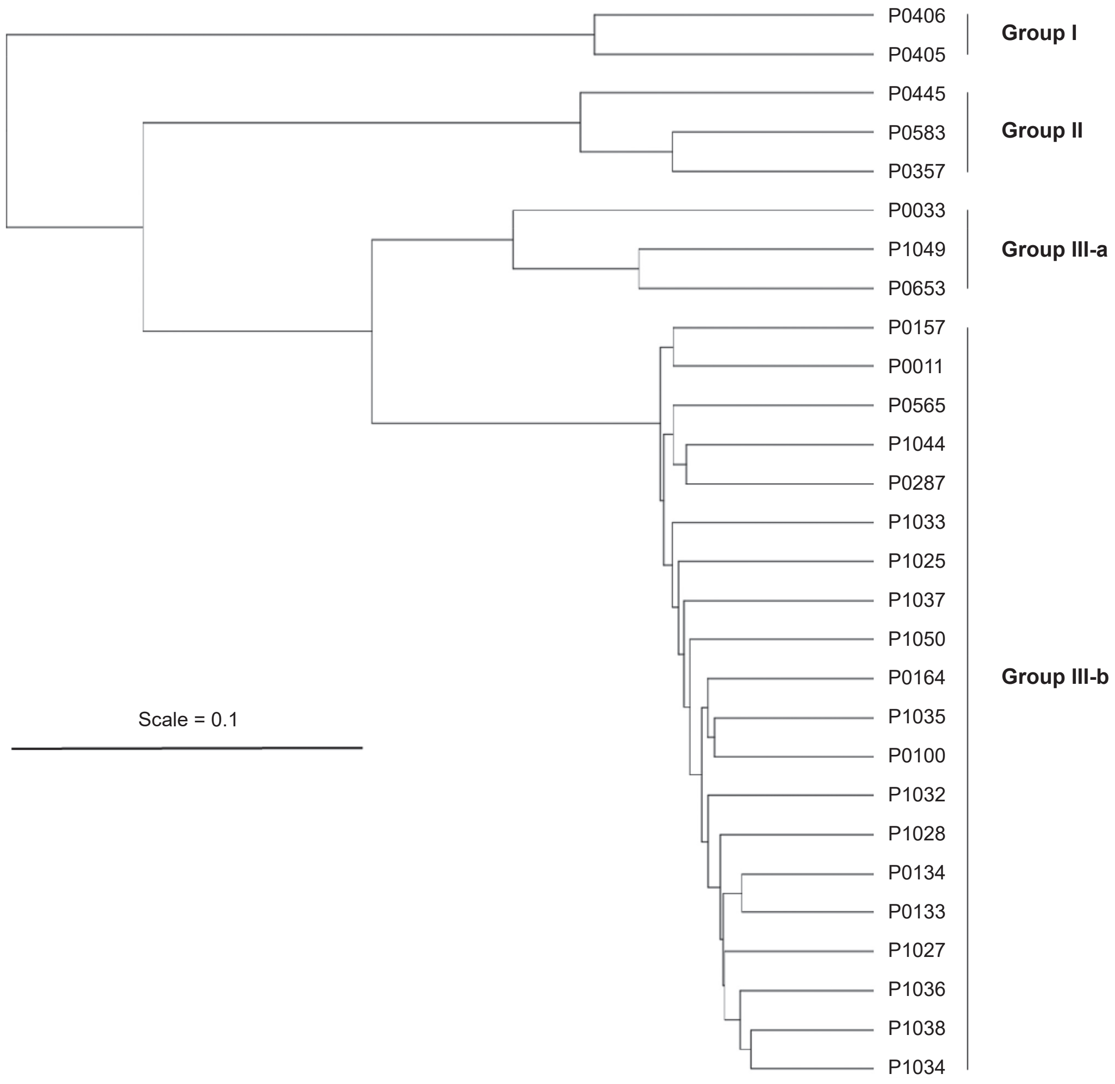

圖 6. 28 個番椒收集系之親緣關係樹。此圖是使用 28 個番椒在 14,230 個單一核苷酸多型性位點的基因型所建構。

Fig. 6. Rooted phylogenetic tree based on the genotypes of 14,230 single nucleotide polymorphisms (SNPs) of the 28 *Capsicum* accessions.

為原始讀序數量的 87–92% 及 43–46%。此顯示在相同初始 gDNA 質量及相同建庫方法下，*Pst*I/*Msp*I 可得到較多成對且比對專一的讀序。

本研究測試的 2 組酵素 *Pst*I/*Msp*I 及 *Pst*I/*Mse*I，其 rare cutter 皆為 *Pst*I，屬甲基化敏感 (methylation-sensitive) 的限制酶，另一端 (common cutter) 則分別為甲基化敏感的 *Msp*I 及甲基化不敏感 (methylation-insensitive) 的 *Mse*I。前人研究顯示甲基化敏感的限制酶能更有效針對富含基因的區段進行酶切，從而降低基因體複雜度 (減少 repeat element) (Fellers 2008; Zhang *et al*. 2010)。Pootakham (2016) 等人針對油棕的基因體研究更實際指出，使用 2 個非甲基化敏感的限制酶時，僅 34% 的 SNP 位在基因上或附近 (5 Kb)，同時使用 2 個甲基化敏感的限制酶時，則有 60–67% 的 SNP

落於基因上或附近。考量 *Pst*I/*Msp*I 組合可更集中定序資源於基因豐富的區域，且不論在建庫資源利用上或是成功對應回參考序列的讀序數量/比例上，*Pst*I/*Msp*I 皆較 *Pst*I/*Mse*I 表現佳，爰此，本研究第二部分將以 *Pst*I/*Msp*I 組合進行番椒種原 ddRAD 建庫，並進行番椒種原遺傳背景分析。

本研究第二部分針對 28 個分屬 5 個馴化種的番椒收集系，以 *Pst*I/*Msp*I 酶切並進行 ddRAD 定序，共得到 850,400 個 SNP，經篩選過後以 14,230 個高品質 SNP 進行後續番椒種原基因體初探。如圖 4，在 28 個番椒收集系中，不同種番椒材料經酶切及定序所得可成對對應到參考序列且只對應到 1 個基因體上的位置讀序數目約占原始讀序數目之75.6–93.9%，大致上以 *C. pubescens* 稍低 (小於 80%)，其餘均高於 80%，說明不同種番椒之間之基因體組成與序列仍具有差異，然而無論有效讀數占比在不同種番椒之間具有差異，後續仍可進行 SNP 分析。另外，相比於其他有關番椒種原 (涵蓋 5 個馴化種) 簡化基因體定序的文獻，Ramchiary & Kole (2019) 使用 14,444 個 SNP、Pereira-Dias *et al*. (2019) 使用 4,083 個 SNP、Potnis *et al*. (2019) 使用90,794個SNP，由於每篇研究使用的建庫方法、定序方法、SNP 篩選條件等皆不同，因此在 SNP 數量上無法進行比較，不過以 SNP 於番椒基因體分布情形來看，本研究得到的 14,230 個高品質 SNP 涵蓋全部染色體，並與預期一致地以染色體兩端基因較密集處較多，染色體中段亦無明顯缺失 SNP 處，因此這些 SNP 應可用於番椒基因體分析。為初步瞭解此些 SNP 用於跨種番椒基因體分析是否合適，因此本研究進一步將此些 SNP 對不同種番椒進行集群分析。

以 28 個分屬 5 個馴化種的番椒收集系之 14,230 個 SNP 進行集群分析，主成分分析圖及樹狀圖皆顯示 *C. baccatum* 及 *C. pubescens* 的品系可各自為一分群或一分支，*C. annuum*、*C. chinense* 及 *C. frutescens* 則先分到同一分支之後，才又將 *C. annuum*、*C. chinense* 及 *C. frutescens* 分開，而 *C. chinense* 及 *C. frutescens* 則較難區分。前人研究以性狀與分子標誌技術將番椒種類分成 3 大群，其中 *C. annuum*、*C. chinense* 及 *C. frutescens* 因親緣性接近，合稱為 *C. annuum* complex，另外 2 種 *C. baccatum* 及 *C. pubescens* 在分類上差異大，故分別以 *C. baccatum* complex 及 *C. pubescens* complex 稱呼 (Walsh & Hoot 2001; García *et al*. 2016; Shiragaki *et al*. 2020)。本研究分群結果與前人研究一致，不過有 2 個品系的分群結果與原先種原資料不符，P0011 (*C. chinense*) 與 P0157 (*C. frutescens*) 被分到 Group III-b (多為 *C. annuum*)，可能原因為已與 *C. annuum* 雜交，或是原先種原資料登記錯誤，實際觀察分類上重要特徵，如：花瓣色、花藥色、萼頸環縮 (caylx annular constriction) 及萼緣齒突 (teeth) 等，亦無法清楚分辨此二品系之種名。本研究 5 個馴化種的集群分析結果與前人研究相符，說明本研究使用的一系列建庫、定序及生物資訊分析等方法可作為跨種番椒種原基因體分析之用。

結論

次世代定序技術已廣泛用於作物基因體研究，亦延伸出許多相關技術，然而每個作物適宜的建庫方法不同，有必要針對目標作物先進行方法可行性測試及比較。本研究目標作物為番椒種原，由於種原涵蓋不同種且通常有收集系內仍有變異的情形，使得複雜度又比常見的雙親本族群更高，此外番椒基因體亦相當大，在定序上須更大量經費，因此本研究以雙重酶切 DNA 定序技術降低番椒基因體複雜度，同時並節省定序資源。本研究比較不同限制酶組合所產生的資料，並用於番椒種原集群分析，結果顯示本研究採用的建庫、定序、生物資訊分析等方法，可將番椒種原正確分類，此說明 *Pst*I/*Msp*I 雙重酶切 DNA 定序方法為跨種番椒基因體分析之良好工具。未來將以此為基礎，針對台灣番椒種原建立其基因體資訊，作為台灣番椒育種的根基。

誌謝

本研究之 ddRAD 建庫方法由國立台灣大

學農藝學系胡凱康老師研究室慷慨傳授，經費由行政院農業委員會 110 農科-4.6.1-農-C9 計畫支持，謹致謝忱。

引用文獻

Altshuler, D., V. J. Pollara, C. R. Cowles, W. J. Van Etten, J. Baldwin, L. Linton, and E. S. Lander. 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407:513–516. doi:10.1038/35035083

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376. doi:10.1371/journal.pone.0003376

Campbell, E. O., B. M. T. Brunet, J. R. Dupuis, and F. A. H. Sperling. 2018. Would an RRS by any other name sound as RAD? Methods Ecol. Evol. 9:1920–1927. doi:10.1111/2041-210X.13038

Cumer, T., C. Pouchon, F. Boyer, G. Yannic, D. Rioux, A. Bonin, and T. Capblancq. 2021. Double-digest RAD-sequencing: Do pre- and post-sequencing protocol parameters impact biological results? Mol. Genet. Genomics 296:457–471. doi:10.1007/s00438-020-01756-9

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158. doi:10.1093/bioinformatics/btr330

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. 2021. Twelve years of SAMtools and BCFtools. Gigascience 10:giab008. doi:10.1093/gigascience/giab008

Du, H., J. Yang, B. Chen, X. Zhang, J. Zhang, K. Yang, S. Geng, and C. Wen. 2019. Target sequencing reveals genetic diversity, population structure, core-SNP markers, and fruit shape-associated loci in pepper varieties. BMC Plant Biol. 19:578. doi:10.1186/s12870-019-2122-2

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379. doi:10.1371/journal.pone.0019379

Fellers, J. P. 2008. Genome filtering using methylation-sensitive restriction enzymes with six base pair recognition sites. Plant Gen. 1:146–152. doi:10.3835/plantgenome2008.05.0245

García, C. C., M. H. J. Barfuss, E. M. Sehr, G. E. Barboza, R. Samuel, E. A. Moscone, and F. Ehrendorfer. 2016. Phylogenetic relationships, diversification and expansion of chili peppers (*Capsicum*, Solanaceae). Ann. Bot. 118:35–51. doi:10.1093/aob/mcw079

Hammer, Ø., D. A. T. Harper, and P. D. Ryan. 2001. PAST: paleontological statistics software package for education and data analysis. Palaeontol. Electron. 4(1):4.

Han, K., H. Y. Lee, N. Y. Ro, O. S. Hur, J. H. Lee, J. K. Kwon, and B. C. Kang. 2018. QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. Plant Biotechnol. J. 16:1546–1558. doi:10.1111/pbi.12894

Hulse-Kemp, A. M., S. Maheshwari, K. Stoffel, T. A. Hill, D. Jaffe, S. R. Williams, N. Weisenfeld, S. Ramakrishnan, V. Kumar, P. Shah, M. C. Schatz, D. M. Church, and A. van Deynze. 2018. Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. Hortic. Res. 5:4. doi:10.1038/s41438-017-0011-0

Kao, C. 2016. Construction of a high density linkage map with double digest RADseq and mapping of restorer-of fertility gene in pepper (*Capsicum annuum* L.). Master Thesis. Department of Agronomy, National Taiwan University. Taipei, Taiwan. 94 pp. (in Chinese with English abstract) doi:10.6342/NTU201610013

Kim, S., M. Park, S. I. Yeom, Y. M. Kim, J. M. Lee, H. A. Lee, E. Seo, J. Choi, K. Cheong, K. T. Kim, K. Jung, G. W. Lee, S. K. Oh, C. Bae, S. B. Kim, H. Y. Lee, S. Y. Kim, M. S. Kim, B. C. Kang, Y. D. Jo, H. B. Yang, H. J. Jeong, W. H. Kang, J. K. Kwon, C. Shin, J. Y. Lim, J. H. Park, J. H. Huh, J. S. Kim, B. D. Kim, O. Cohen, I. Paran, M. C. Suh, S. B. Lee, Y. K Kim, Y. Shin, S. J. Noh, J. Park, Y. S. Seo, S. Y. Kwon, H. A Kim, J. M. Park, H. J. Kim, S. B. Choi, P. W. Bosland, G. Reeves, S. H. Jo, B. W. Lee, H. T. Cho, H. S. Choi, M. S. Lee, Y. Yu, Y. D. Choi, B. S. Park, A. van Deynze, H. Ashrafi, T. Hill, W. T. Kim, H. S. Pai, H. K. Ahn, I. Yeam, J. J. Giovannoni, J. K. C. Rose, I. Sørensen, S. J. Lee, R. W. Kim, I. Y. Choi, B. S. Choi, J. S. Lim, Y. H. Lee, and D. Choi. 2014. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat. Genet. 46:270–278. doi:10.1038/ng.2877

Kim S., J. Park, S. I. Yeom, Y. M. Kim, E. Seo, K. T. Kim, M. S. Kim, J. M. Lee, K. Cheong, H. S. Shin, S. B. Kim, K. Han, J. Lee, M. Park, H. A. Lee, H. Y. Lee, Y. Lee, S. Oh, J. H. Lee, E. Choi, E. Choi, S. E. Lee, J. Jeon, H. Kim, G. Choi, H. Song, J. Lee, S. C. Lee, J. K. Kwon, H. Y. Lee, N. Koo, Y. Hong, R. W. Kim, W. H. Kang, J. H. Huh, B. C. Kang, T. J. Yang,

Y. H. Lee, J. L. Bennetzen, and D. Choi. 2017. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biol. 18:210. doi:10.1186/s13059-017-1341-9

Lepais, O. and J. T. Weir. 2014. SimRAD: An R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol. Ecol. Resour. 14:1314–1321. doi:10.1111/1755-0998.12273

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303:3997. doi:10.48550/arXiv.1303.3997

Nimmakayala, P., V. L. Abburi, T. Saminathan, A. Almeida, B. Davenport, J. Davidson, C. V. C. M. Reddy, G. Hankins, A. Ebert, D. Choi, J. Stommel, and U. K. Reddy. 2016. Genome-wide divergence and linkage disequilibrium analyses for *Capsicum baccatum* revealed by genome-anchored single nucleotide polymorphisms. Front. Plant Sci. 7:1646. doi:10.3389/fpls.2016.01646

Pereira-Dias, L., S. Vilanova, A. Fita, J. Prohens, and A. Rodríguez-Burruezo. 2019. Genetic diversity, population structure, and relationships in a collection of pepper (*Capsicum* spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). Hortic. Res. 6:54. doi:10.1038/s41438-019-0132-8

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135. doi:10.1371/journal.pone.0037135

Pootakham, W., C. Sonthirod, C. Naktang, N. Jomchai, D. Sangsrakru, and S. Tangphatsornruang. 2016. Effects of methylation-sensitive enzymes on the enrichment of genic SNPs and the degree of genome complexity reduction in a two-enzyme genotyping-by-sequencing (GBS) approach: A case study in oil palm (*Elaeis guineensis*). Mol. Breed. 36:154. doi:10.1007/s11032-016-0572-x

Potnis, N., S. E. Branham, J. B. Jones, and W. P. Wechter. 2019. Genome-wide association study of resistance to *Xanthomonas gardneri* in the USDA pepper (*Capsicum*) collection. Phytopathology 109:1217–1225. doi:10.1094/PHYTO-06-18-0211-R

Qin, C., C. Yu, Y. Shen, X. Fang, L. Chen, J. Min, J. Cheng, S. Zhao, M. Xu, Y. Luo, Y. Yang, Z. Wu, L. Mao, H. Wu, C. Ling-Hu, H. Zhou, H. Lin, S. González-Morales, D. L. Trejo-Saavedra, H. Tian, X. Tang, M. Zhao, Z. Huang, A. Zhou, X. Yao, J. Cui, W. Li, Z. Chen, Y. Feng, Y. Niu, S. Bi, X. Yang, W. Li, H. Cai, X. Luo, S. Montes-Hernández, M. A. Leyva-González, Z. Xiong, X. He, L. Bai, S. Tan, X. Tang, D. Liu, J. Liu, S. Zhang, M. Chen, L. Zhang, L. Zhang, Y. Zhang, W. Liao, Y. Zhang, M. Wang, X. Lv, B. Wen, H. Liu, H. Luan, Y. Zhang, S. Yang, X. Wang, J. Xu, X. Li, S. Li, J. Wang, A. Palloix, P. W. Bosland, Y. Li, A. Krogh, R. F. Rivera-Bustamante, L. Herrera-Estrella, Y. Yin, J. Yu, K. Hu, and Z. Zhang. 2014. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proc. Natl. Acad. Sci. U.S.A. 111:5135–5140. doi:10.1073/pnas.1400975111

Ramchiary, N. and C. Kole. 2019. The Capsicum Genome. Springer Cham. Aargau, Switzerland. 232 pp. doi:10.1007/978-3-319-97217-6

Shiragaki, K., S. Yokoi., and T. Tezuka. 2020. Phylogenetic analysis and molecular diversity of *Capsicum* based on rDNA-ITS region. Horticulturae 6:87. doi:10.3390/horticulturae6040087

Taitano, N., V. Bernau, L. Jardón-Barbolla, B. Leckie, M. Mazourek, K. Mercer, L. McHale, A. Michel, D. Baumler, M. Kantar, and E. van der Knaap. 2019. Genome-wide genotyping of a novel mexican chile pepper collection illuminates the history of landrace differentiation after *Capsicum annuum* L. domestication. Evol. Appl. 12:78–92. doi:10.1111/eva.12651

Thorvaldsdóttir, H., J. T. Robinson, and J. P. Mesirov. 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief. Bioinform. 14:178–192. doi:10.1093/bib/bbs017

Walsh, B. M. and S. B. Hoot. 2001. Phylogenetic relationships of *Capsicum* (Solanaceae) using DNA sequences from two noncoding regions: The chloroplast *atpB-rbcL* spacer region and nuclear *waxy* introns. Int. J. Plant Sci. 162:1409–1418. doi:10.1086/323273

Yamazaki, A., K. Shirasawa, and M. Hosokawa. 2020. Transgressive segregation and gene regions controlling thermotolerance of fruit set and pollen germination in *Capsicum chinense*. Euphytica 216:179. doi:10.1007/s10681-020-02712-9

Zhang, M., J. N. Kimatu, K. Xu, and B. Liu. 2010. DNA cytosine methylation in plant development. J. Genet. Genom. 37:1–12. doi:10.1016/S1673-8527(09)60020-5

The Application of Double-Digest Restriction-Site Associated DNA Sequencing Technology in Pepper Germplasm

Ssu-Yu Lin[1], Jau-Yueh Wang[1], Cheng-Bin Li[1], Han-Wei Chen[1], Da-Gin Lin[2], and Yuan-Kai Tu[1,*]

Abstract

Lin, S. Y., J. Y. Wang, C. B. Li, H. W. Chen, D. G. Lin, and Y. K Tu. 2023. The application of double-digest restriction-site associated DNA sequencing technology in pepper germplasm. J. Taiwan Agric. Res. 72(2):97–112.

Next-generation sequencing technologies lead to a new era of crop genetic research. To lower sequencing costs for building a single nucleotide polymorphism (SNP) map of a genome, the reduced representation genome sequencing technology was used to sequence the flanking DNA sequences of selected restriction enzyme sites of DNA samples. However, appropriate parameters for constructing the reduced representation sequencing library, such as the choice of restriction enzyme and sequencing depth, differ in target crops and population composition. As a result, methodological adjustments toward target crops and population composition are vital to maximizing the efficiency of sequencing resources. In this study, we presented a tailored double-digest restriction site-associated DNA sequencing (ddRAD-seq) genotyping method for pepper germplasm. In the first part of this work, the performance of two restriction enzyme pairs- *Pst*I/*Msp*I and *Pst*I/*Mse*I were assessed, and results of *in silico* simulation showed 2.01% and 0.34% fragments of the total digested fragment were sequenced, respectively. According to sequencing results, more uniquely mapped and adequately paired reads (19–28 M) were observed in *Pst*I/*Msp*I pair than in *Pst*I/*Mse*I (15–17 M). Therefore, in terms of resource utilization in library preparation, the enzyme pair *Pst*I/*Msp*I performed better than *Pst*I/*Mse*I. In addition, the number or ratio of reads that came from *Pst*I/*Msp*I digestion was more efficiently mapped to the reference genome. As for the second part of this work, the SNP maps of 28 pepper accessions belonging to 5 domestic species were constructed using the ddRAD-seq genotyping technique with *Pst*I/*Msp*I restriction enzyme combination. A total of 14,230 high-quality SNP were obtained and distributed evenly on the 12 chromosomes of pepper. Furthermore, as expected, the 28 pepper accessions were clustered into proper phylogenetic groups based on the principle component analysis and the dendrogram analysis. In summary, the ddRAD-seq genotyping technique was tailored to construct the SNP maps of the pepper germplasm in this study.

Key words: Reduced genome sequencing, Double-digest restriction enzyme, *Capsicum*, Genome analysis.

Received: August 12, 2022; Accepted: January 7, 2023.
[*] Corresponding author, e-mail: yktu@tari.gov.tw
[1] Assistant Research Fellows, Biotechnology Division, Taiwan Agricultural Research Institute, Taichung City, Taiwan, ROC.
[2] Contract Associate Research Fellow, Biotechnology Division, Taiwan Agricultural Research Institute, Taichung City, Taiwan, ROC.